

NAME: \_\_\_\_\_

142  
Ch. 1 review~~000000~~**MULTIPLE CHOICE:**

2. Were the extinctions that occurred in the last ice age more frequent among species of animals with large body sizes? A researcher gathers data on the average body mass (in kilograms) of all species known to have existed at that time. These measurements are values of

- (a) a categorical variable.
- ☒ (b) a quantitative variable.
- (c) an invalid variable.
- (d) a margin of error.

4. An example of a categorical variable is

- (a) the name of the college a student attends.
- (b) a student's weight in kilograms.
- (c) a student's class rank, such as 25th out of 364.
- (d) a student's sex (male or female).
- ☒ (e) Both (a) and (d).

The stock market did well during the 1990s. Here are the percent total returns (change in price plus dividends paid) for the Standard & Poor's 500 stock index:

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Return	31.7	-3.1	30.5	7.6	10.1	1.3	37.6	23.0	33.4	28.6

The next five questions are related to this situation.

5. The median **return** during this period is

- (a) 5.5
- (b) 20.07
- (c) 23.0
- ☒ (d) 25.8
- (e) 28.6

6. The third quartile of these **returns** is

- (a) 7.6
- (b) 30.5
- (c) 31.1
- ☒ (d) 31.7
- (e) 33.4

7. The mean **return** is

- ☒ (a) 20.07
- (b) 20.69
- (c) 22.3
- (d) 25.8
- (e) 33.4

8. The standard deviation of the **returns** is

- (a) 13.75
- (b) 13.98
- ☒ (c) 14.74
- (d) 20.07
- (e) 25.8

18. A well-drawn histogram should have
- (a) bars all the same width
  - (b) no space between bars (unless a class has no observations)
  - (c) a clearly marked vertical scale
  - ☒ (d) all of these

Here is a table of the undergraduate enrollment at a large state university, broken down by class:

Class	Count of students	Percent of Students
Freshman	8,248	26.8%
Sophomore	8,073	26.2%
Junior	7,001	22.8%
Senior	6,904	22.4%
Non-degree	535	1.7%
Total	30,761	100%

21. To make a correct graph of the distribution of students by class, you could use
- (a) a bar graph.
  - (b) a pie chart.
  - (c) a histogram.
  - (d) all of (a), (b), and (c).
  - ☒ (e) (a) or (b), but not (c).

42. For a distribution that is skewed to the right, usually
- ☒ (a) the mean will be larger than the median
  - (b) the median will be larger than the mean
  - (c) the first quartile will be larger than the third quartile
  - (d) the standard deviation will be negative
  - (e) the minimum will be larger than the maximum

43. Here is a set of data: 1300, 18, 25, 19, -7, 24. Which observation is the outlier?

- ☒ (a) 1300
- (b) 25
- (c) 19
- (d) -7
- (e) 24

45. You have calculated that the value of the standard deviation for a set of data is -2.5. You can conclude that

- (a) the data are skewed to the left
- (b) the data are skewed to the right
- (c) there is very little variation in the data
- (d) all of the observations in the data set are negative
- ☒ (e) you made an error in your calculations

55. You calculate that the standard deviation of a set of observations is 0. This tells you that

- (a) you made an arithmetic mistake.
- ☒ (b) all the observations have the same value.
- (c) there is no straight-line association.
- (d) the mean must also be 0.
- (e) Both (b) and (d) are true.

69. The five-number summary of a distribution consists of

- (a) mean, median, standard deviation, and two quartiles.
- (b) minimum, maximum, mean, median, and standard deviation.
- ☒ (c) minimum, maximum, median, and two quartiles
- (d) mean, standard deviation, correlation, and two quartiles

72. The standard deviation is a measure of

- (a) the center of a distribution
- ☒ (b) the variability of a distribution
- (c) the association between two variables
- (d) the standardized value of a variable

100. The box in the center of a boxplot marks

- (a) the full range covered by the data
- ☒ (b) the range covered by the middle half of the data
- (c) the range covered by the middle three-quarters of the data
- (d) the span one standard deviation on each side of the mean
- (e) the span two standard deviations on each side of the mean

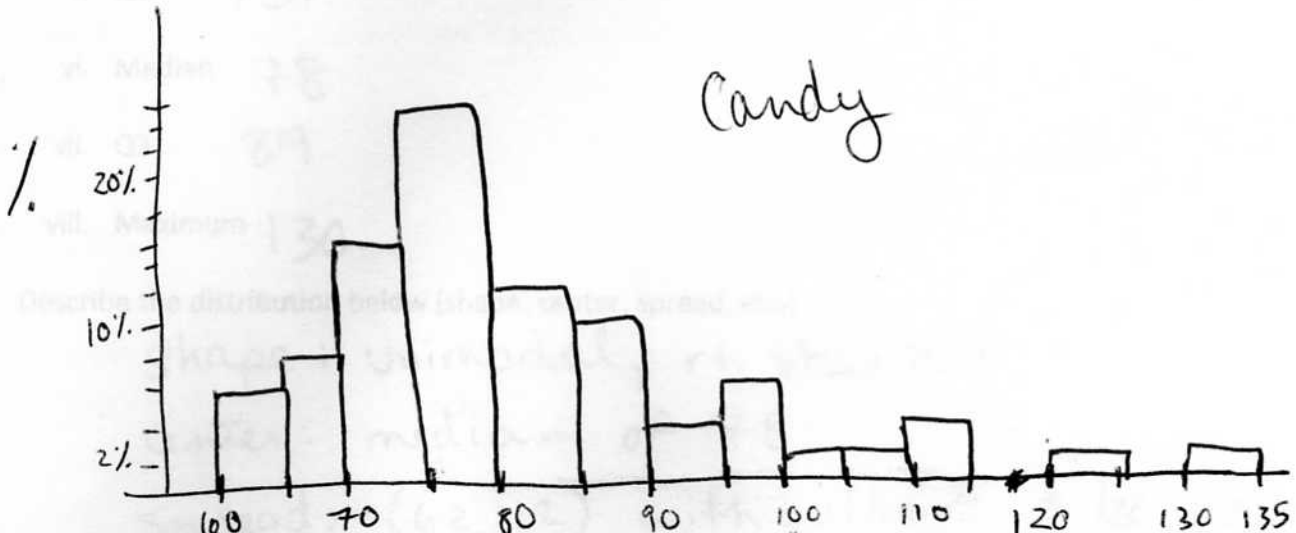
**FREE RESPONSE:**

- 1) The list `CANDY` is a list of the number of pieces of Halloween candy that 50 randomly selected children got on Halloween night. (data is below)

62	66	72	75	76	78	81	86	91	106
62	67	73	75	77	78	81	88	95	111
63	70	73	76	77	79	82	89	96	112
65	71	73	76	77	80	82	89	99	120
66	72	74	76	78	80	85	91	100	130

$n=50$

- a. Create a relative frequency histogram of the data below. Be sure to label all your axes!



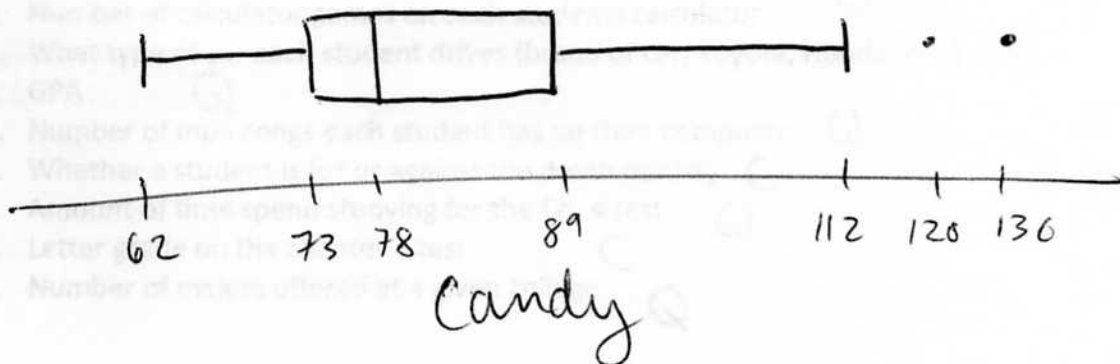
- b. Based on the shape of the distribution, what measure of center would you report (mean or median)?

right skewed  $\Rightarrow$  median

- c. Based on your choice in part b, what measure of spread would you report (std. deviation or 5# summary)?

5 # summary (because we chose median)

- d. Create a boxplot of the data below.



e. Calculate the following statistics

i. Mean 82.02

ii. Std. dev 14.78

iii. Number of observations 50

iv. Min 62

v. Q1 73

vi. Median 78

vii. Q3 89

viii. Maximum 130

f. Describe the distribution below (shape, center, spread, etc.)

shape: unimodal, rt. skewed

center: median of 78

spread: (62, 112) with outliers @ 120, 130

g. Test for outliers using the  $1.5 \times \text{IQR}$  test

$$\text{IQR} = 16$$

$$1.5 \times \text{IQR} = 24$$

$$Q_3 + 24 = 113$$

$$Q_1 - 24 = 49$$

Outliers

120

130

anything above 113 ~~or~~ below 49 is an outlier

2) Identify the following variables as categorical or quantitative

a. Number of calculator games on each student's calculator Q

b. What type of car each student drives (brand of car, Toyota, Honda, etc.) C

c. GPA Q

d. Number of mp3 songs each student has on their computer Q

e. Whether a student is for or against the death penalty C

f. Amount of time spent studying for the Ch. 4 test Q

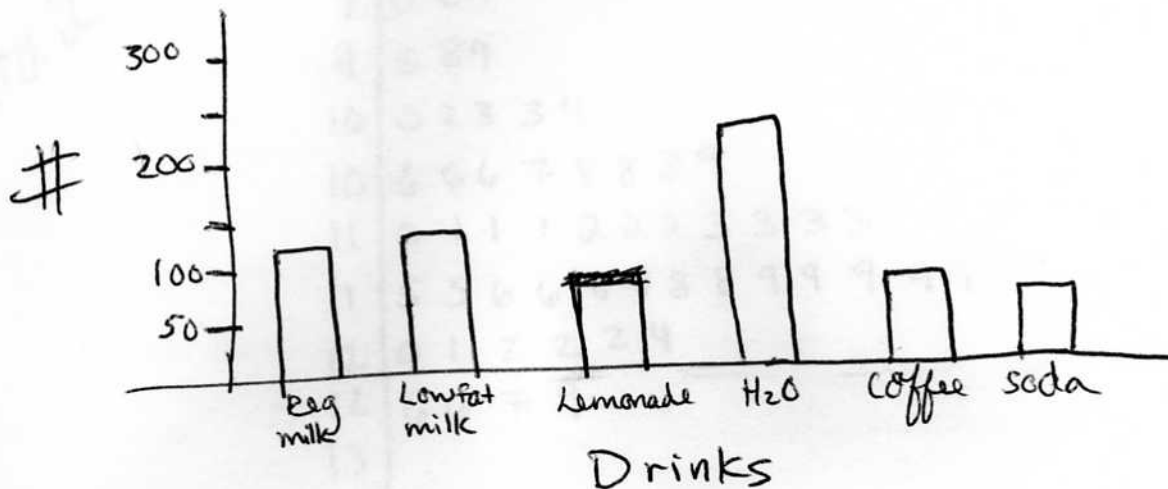
g. Letter grade on the chapter 4 test C

h. Number of majors offered at a given college Q

- 3) The following are the amount of students who buy each of the different drinks offered in the cafeteria during a 5-day long observational study, at a small school.

Regular Milk	120	15%
Lowfat Milk	136	17%
Lemonade	100	12%
Water	250	31%
Coffee	110	14%
Fountain soda	90	11%
<u>n = 806</u>		<u>100%</u>

- a. Create a bar graph of the data (use frequency on the y-axis)



- b. Create a pie chart of the data



- 4) Given the following info about the mean and median, decide the shape of the distribution of the data, and pick a measure of center to report.

a. Mean = 170 Median = 80

Shape = rt. skew Center = M

b. Mean = 100 Median = 180

Shape = left skew Center = M

c. Mean = 79 Median = 85

Shape = roughly symmetric Center =  $\bar{X}$

5) Using the list DATA2 (below), create a stemplot below. You will want to split your stems. How many times you split them is up to you.

82	98	106	108	112	115	119	122	129	108	111	118
85	99	106	109	112	116	119	122	91	108	112	118
89	100	106	110	113	116	119	122	95	126	113	
90	102	107	111	113	118	119	124	103	127	115	
90	103	108	111	113	118	119	126	104	120	121	

Data 2

```

8 | 2
8 | 5 9
9 | 0 0 1
9 | 5 8 9
10 | 0 2 3 3 4
10 | 6 6 6 7 8 8 8 9
11 | 0 1 1 1 2 2 2 3 3 3 3
11 | 5 5 6 6 8 8 8 8 9 9 9 9 9
12 | 0 1 2 2 2 4
12 | 6 6 7 9
13 |

```

a. Describe the distribution (shape, center, spread)

- left skewed
- center ~ 110 stem  
(median of 112)
- spread (82, 129)



b. Calculate the following statistics

i. Mean 110.46

ii. Std. dev 10.94

iii. Number of observations 57

iv. Min 82

v. 1<sup>st</sup> quartile 105

vi. Median 112

vii. 3<sup>rd</sup> quartile 119

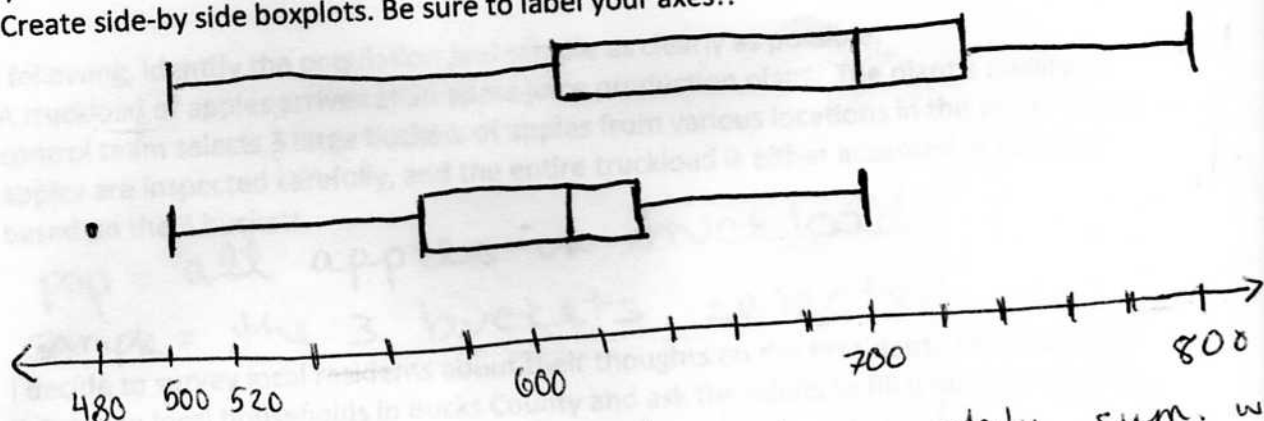
viii. Maximum 129

6) Use the data for the male and female SAT Math scores (in lists  $\text{SATMM}$  for males and  $\text{SATMF}$  for females).

a. Create side-by-side boxplots. Be sure to label your axes!!

Female

male



b. Compare the two plots.

Shapes: females left skewed, males roughly sym. w outlier @ 480.

Center: female median of 700 higher than male of 600

Spread: females more spread (500, 800) than male (480, 700)

c. Where did the top 25% of males score above? How about the top 25% of females?

630

730

d. Where did the middle 50% of males score (list the range)? How about the middle 50% of females?

(575, 630)

(610, 730)



- 7) The following is info on hotels in the area.

Hotel	Pool	exercise room	internet price	restaurants	room service	room rate
Comfort Inn	out	y	0	1	y	\$149
Fairfield Inn	in	y	0	1	n	\$119
Baymont Inn	out	y	0	1	y	\$60
Chase Hotel	out	n	15	0	n	\$139
Courtyard Marriott	in	y	0	1	dinner	\$114
Hilton	in	y	10	2	y	\$156
Best Western	in	y	9.95	2	y	\$145

- a. Identify the individuals

Hotels

- b. Identify the variables measured and label each as categorical or quantitative

Pool - C

Ex. room - C

Internet - Q

restaurants - Q

room serv. - C

room rate - Q

- 8) For the following, identify the population and sample as clearly as possible:

- a. A truckload of apples arrives at an apple juice production plant. The plant's quality control team selects 3 large buckets of apples from various locations in the truck. These apples are inspected carefully, and the entire truckload is either accepted or rejected based on the 3 buckets.

pop = all apples in truckload

sample = the 3 buckets selected

- b. I decide to survey local residents about their thoughts on the President. I send out 100 surveys to local households in Bucks County and ask the adults to fill it out and send it back.

pop = Bucks County households

sample = 100 selected households

- c. A study on cancer patients took a random sample of 80 patients from the lung cancer ward at Temple University Hospital. They then tested their reactions to a new drug to see if there were any improvements on their conditions.

pop = all lung cancer patients

sample = 80 selected patients

9) Identify the following as an experiment or an observational study:

- a. Before a new variety of frozen muffins is put on the market, it is subjected to extensive taste testing. People are asked to taste the new muffin and a competing brand and say which one they prefer.

expt.

- b. A teacher wants to know about how her students spend their time after school. She passes out a survey that asks numerous questions about this topic. The survey is anonymous.

obs. study

- c. Teenage drivers have long been blamed for being bad drivers, and prone to accidents. To decide whether this is a fair statement, a teenage statistics student goes to a local police precinct and obtains a list of all accidents in the past 3 years. She then selects 100 of these accidents and looks to see the age of the person at fault (teenage or adult).

obs. study

- d. There has long been a debate about whether whitening toothpaste does as good of a job as whitening strips. A researcher has 20 people use whitening strips for 5 days and another 20 people use whitening toothpaste for 30 days, and then compares their beginning tooth shade to their new tooth shade.

expt.

10) What is a distribution of a variable?

a picture that shows the values of the variable & how often each occurs. Ex: histogram, stemplot, bar graph

11) What types of values can standard deviation take?

$$s \geq 0$$

12) If I have a standard deviation of 0, what does this mean?

all values are same

13) What are quartiles? How do they divide up the data?

divide data into 4 equal sections in numerical order

14) What is the second quartile?

median

15) What percentage of the data is above Q3? 25%.

16) What percentage of the data is below Q1? 25%.

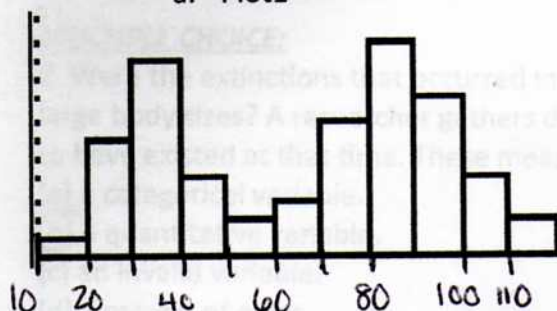
17) What percentage of the data is inside the box of a boxplot? 50%.

18) What is a census?

trying to gather info from an entire population

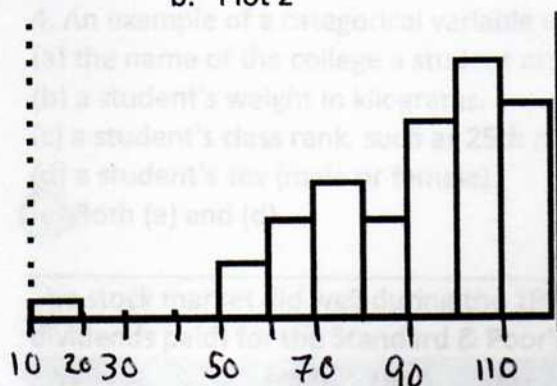
19) Describe the following distributions (shape, center, spread, etc.). All plots begin at 10, and the bar width is 10 units.

a. Plot 1



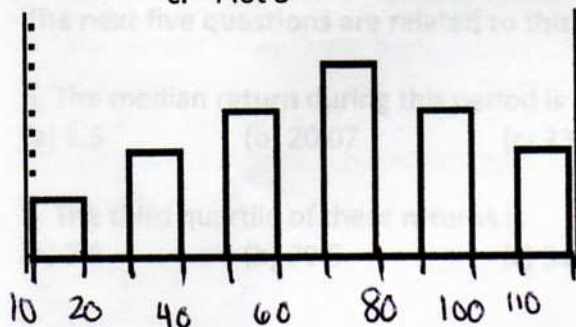
- bimodal
- symmetric
- center  $\sim 60$
- spread (10, 110)

b. Plot 2



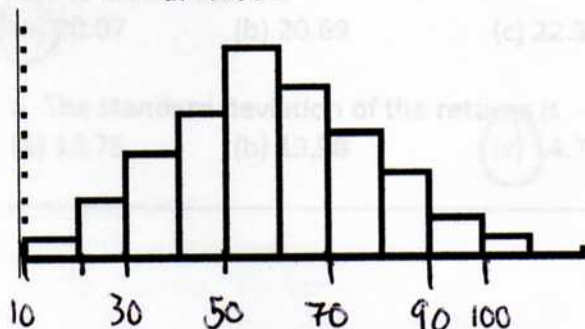
- unimodal
- left skewed
- poss. outlier @ 10
- range (50, 110) + outlier @ 10
- center  $\sim 80$

c. Plot 3



- granularity
- unimodal
- symmetric
- range (10, 110)
- center  $\sim 70$

d. Plot 4



- symmetric
- unimodal
- center  $\sim 50$
- spread (10, 100)